４．学習問題における従来研究と統計的決定理論に基づく本研究の位置付け

本章では文書分類問題，既存名詞シソーラスへの未知語分類問題，未知情報を伴うマルコフ決定過程問題の従来研究について述べるとともに，従来研究と本研究の目的の違いを述べることによって，本研究の位置付けを明確にする．

４．１．分類問題の従来研究

４．１．１．分類問題の概要

分類問題とは分類すべき分類対象（文書分類問題における文書，既存名詞シソーラスへの未知語分類問題における未知語等）とその分類対象に付随する分類するための手がかりとなるキー情報（文書分類問題における文書中のキーワード，既存名詞シソーラスへの未知語分類問題における未知語と共起した動詞等）を入力として受け取ったもとで，キー情報を手がかりとして分類対象をクラスに分類する問題である．一般的に分類問題というと，クラスの集合が既知で学習データ（正解のクラスとキー情報の組）が与えられたもとで新規の分類対象を分類対象に付随するキー情報を手がかりとしてクラスに分類する問題と，クラスの集合が未知で学習データが与えられないもとで分類したい分類対象の集合を各分類対象に付随するキー情報を手がかりとして何らかの基準のもとで分類する問題（クラスタリング問題）の２つに大別される．本論文ではこれ以降，分類問題という言葉を前者のクラス集合既知の問題のみを指して使うこととする．

ここで，簡単な定義を行う．クラスをc_i，$C=\{c_1,c_2,\cdots,c_{|C|}\}$として，分類の手がかりとなるキー情報を$key_j$，$KEY=\{key_1,key_2,\cdots,key_{|KEY|}\}$とし，キー情報$key_j$はクラス$c_i$から生起するものとする．クラスとキー情報系列の$n$個の 2 項組による学習データを$(x,y)^n$，$x\in C$，$y\in KEY$，新規に分類したい分類対象のキー情報を$y'$，$y'\in KEY$，キー情報$y'$はクラス$x'$，$x'\in C$から生起したものとする．分類問題は学習データ$(x,y)^n$と新規に分類したい分類対象のキー情報$y'$を受け取ったもとで，クラス$x'$を推定する問題である．

分類問題の解決方法である分類方法は従来から距離を用いる分類方法と確率モデルを用いる分類方法に大別される．前者の距離を用いる分類方法では，新規に分類したい分類対象と各クラス間の何らかの距離を定義し，その距離が最小となるクラスに新規の分類対象を分類することを基本的な考え方としている．この種の方法は適用可能な範囲が非常に広く，かつ実装が容易であり，自然言語処理におけるさまざまな言語的性質などを加味した調整も容易であるため，とても多くの応用的な研究がなされると共に，実際に実用化されている事例も多い．ただし，その分類精度にはあまり理論的な保証はない．以下，距離を

用いる分類方法をいくつか紹介する．[1]

距離を用いる分類方法にはテンプレートマッチング法・k最近傍法等がある．テンプレートマッチング法は，学習データ中のキー情報をクラスごとにまとめて一つの代表ベクトル（テンプレート）を作成し，新規に分類したい分類対象のキー情報から作成したベクトルと各クラスの代表ベクトル間の距離（あるいは類似度）を測定して最も近いクラスに分類する方法である．この方法は直感的に大変分り易いという長所があるが，分類の精度は各クラスの代表ベクトルの作成方法やベクトル間の距離の決め方次第である．k最近傍法は，学習データをすべて記憶しておき，新規に分類したい分離対象のキー情報と近いk個の学習データを抽出し，抽出されたk個の学習データの中で多数決を行い，最も多いクラスに新規の分類対象を分類する方法である．学習データをすべて記憶しているのでデータのばらつきにも柔軟に対応できるという長所があるが，分類精度はテンプレートマッチング法同様に距離の決め方次第である．

距離を用いる分類方法の中で代表的な方法にテンプレートマッチング法の一種のベクトル空間法がある．ベクトル空間法は実装等が容易であることに加え，特に計算量が小さいという長所を有すために自然言語処理の分野等で最もよく研究および実用化されている．例えば自然言語処理の分野でベクトル空間法が利用されている例を列挙してみると，情報検索問題，情報フィルタリング問題，機械翻訳問題等が挙げられる．以下，簡単に各問題について紹介する．情報検索問題はキーワードによる検索問題と自然文による検索問題（質問応答問題とも呼ぶ．）に大別される．前者は多くの検索エンジンや検索機能付の文書ＤＢなどで利用されており，エンドユーザが入力したキーワード系列（キー情報系列）がＤＢ内のどの文書（情報）と近いかを判断することによって検索結果を求めている．後者の質問応答問題では，エンドユーザによる自然文での質問から質問の種類を示すキー情報と質問の具体的な対象を示すキー情報を抽出した上で，対象を示すキー情報を用いてキーワードによる検索を実施し，該当する質問の種類に即した回答文を作成している．情報フィルタリング問題はある設定された条件（キーワード）に関して情報検索問題と同様の検索処理を日々行って，新規の情報があればエンドユーザに情報を提供することを実施していると解釈できる．機械翻訳問題では語義の多義を解消する際にキー情報（動詞等）との共起頻度を利用してシソーラス上で最も近い語義（クラス）を見つける等の処理が実施されている．本研究では分類問題の従来研究の代表的な方法の一つとして特にこのベクトル空間法を後で取り上げる．

確率モデルを用いる分類方法では，分類対象の生起（クラスおよびキー情報の生起）の仕方に確率モデルを仮定し，その上で分類を誤る確率である誤り率を評価尺度として導入し，誤り率をなるべく小さくするように分類を行うことを基本的な考え方としている．確率分布を仮定しているので，誤り率という評価尺度のもとで理論的な最適性などを保証することができる長所がある一方で，確率分布を仮定できない問題には適用できないという短所がある．確率モデルを用いる分類方法の中で自然言語処理等の分野で最もよく研究さ

れている代表的な方法が Naive-Bayes 法である．本研究では分類問題の従来研究の代表的な分類方法の 1 つとして特にこの Naive-Bayes 法を後で取り上げる．なお，本研究で提案する提案アルゴリズムもこの確率モデルを用いる分類方法の一種である．

分類問題にはさまざまな問題が包含されるが，本研究ではその中から自然言語処理の分野における文書分類問題と既存名詞シソーラスへの未知語分類問題を研究対象とする．

４．１．２．文書分類問題における従来研究

以下，本研究では文書の例として電報を用いるが，一般的な文書分類問題と特に違いはない．文書分類問題について述べる前に，まず，いくつかの定義を行う．c_i は電報が内容に応じて分類される，「結婚」，「クリスマス」等のクラスを示し，$c_i \in C$ で，C は電報のクラスの集合，$C = \{c_1, c_2, \cdots, c_{|C|}\}$ である．なお，$|\cdot|$ は集合の要素数を示す．前提として，一つの電報は必ず一つのクラスのみに分類されるものとする．key_j は電報の電報文中に出現するキー情報（キーワード）を示し，$key_j \in KEY$ で，KEY はキー情報の集合，$KEY = \{key_1, key_2, \cdots, key_{|KEY|}\}$ である．

文書分類問題とは，既に各クラス c_i に分類されている学習用の電報 doc_{l_i} を用いて学習し，新規に分類したい未知の電報 doc_u をいずれかのクラス c_i に分類する問題である．

学習用の電報 doc_{l_i}（学習データ）は，doc_{l_i} が分類されているクラス x_i，$x_i \in C$ と doc_{l_i} の電報文に含まれるキー情報の系列 y^{N_i} の 2 項組 (x_i, y^{N_i}) で表現される．なお，N_i はキー情報系列 y^{N_i} の長さ（doc_{l_i} に含まれるキー情報の延べ数に相当する．）を示し，y^{N_i} は $y_{i,1}y_{i,2}\cdots y_{i,N_i}$ と同一で，$y_{i,j}$，$y_{i,j} \in KEY$ は doc_{l_i} に含まれるキー情報の系列中の j 番目に並んでいるキー情報を示す．学習用の電報 doc_{l_i} は n 個与えられ，n 個の学習用の電報全体は，長さ n の電報の系列 ${doc_{l_i}}^n$ として表現される．また，${doc_{l_i}}^n$ は $doc_{l_1}doc_{l_2}\cdots doc_{l_n}$ や $(x_i, y^{N_i})^n$ や $(x, y^N)^n$ のように表現されることもある．

新規に分類したい未知の電報 doc_u は，doc_u が分類されるべき真のクラス x'，$x' \in C$ と doc_u の電報文に含まれるキー情報の系列 $y'^{N'}$ の 2 項組 $(x', y'^{N'})$ で表現される．なお，N' はキー情報系列 $y'^{N'}$ の長さ（未知の電報 doc_u に含まれるキー情報の延べ数に相当する．）を

示し，$y'^{N'}$ は $y'_1 y'_2 \cdots y'_{N'}$ と同一である．また，実際に与えられるのはキー情報の系列 $y'^{N'}$ のみで，真のクラス x' は未知である．

すなわち，電報の文書分類問題とは，新規に分類したい未知の電報 doc_u のキー情報系列 $y'^{N'}$ を与えられたもとで，doc_u が分類されるべき真のクラス x' を推定する問題として解釈できる．

文書分類問題に関しては，従来からさまざまな研究が行われているが大別すると距離を用いた分類方法と確率モデルを用いた分類方法に分けられる．最もよく研究および実用化されている基本的な従来方法として，距離を用いた分類方法の代表的な方法であるベクトル空間法に基づく文書分類方法と，確率モデルを用いた分類方法の代表的な方法である Naive-Bayes 法に基づく文書分類方法を以下で説明する．

情報検索の分野で広く用いられているベクトル空間法を文書分類に適用した文書分類方法がいくつか提案されており[12][33][40][43]，文書分類に適した様々な提案が盛り込まれたアルゴリズムが提案されている．しかし，本研究ではベクトル空間法の基本的な性質に着目したいので，ここでは，最も単純な TF･IDF 法を加味した基本的なベクトル空間法に基づく文書分類方法を紹介する．なお，以下では単にベクトル空間法に基づく文書分類方法と呼ぶ．

ベクトル空間法に基づく文書分類方法では，次式によって分類するクラスが決定される．なお，$d_{Vec}\left(y'^{N'}\right)$ は未知の電報 doc_u のキー情報系列 $y'^{N'}$ を引数にとり，ベクトル空間法に基づいて未知の電報を分類するクラスを決定する決定関数である．

$$d_{Vec}\left(y'^{N'}\right) = \arg\max_{\hat{x}' \in C} \cos\left(V(\hat{x}'), V(doc_u)\right) = \arg\max_{\hat{x}' \in C} \frac{V(\hat{x}') \cdot V(doc_u)}{\|V(\hat{x}')\| \|V(doc_u)\|}, \tag{4.1}$$

ただし，

$$V(\hat{x}') = \left(F\left((\hat{x}', key_1) \middle| (x, y^N)^n \right) \log \frac{|C|}{A(key_1)}, F\left((\hat{x}', key_2) \middle| (x, y^N)^n \right) \log \frac{|C|}{A(key_2)}, \right.$$
$$\left. \cdots, F\left((\hat{x}', key_{|KEY|}) \middle| (x, y^N)^n \right) \log \frac{|C|}{A(key_{|KEY|})} \right), \tag{4.2}$$

$$V(doc_u) = \left(F\left(key_1 \middle| y'^{N'}\right) \log \frac{|C|}{A(key_1)}, F\left(key_2 \middle| y'^{N'}\right) \log \frac{|C|}{A(key_2)}, \right.$$
$$\left. \cdots, F\left(key_{|KEY|} \middle| y'^{N'}\right) \log \frac{|C|}{A(key_{|KEY|})} \right), \tag{4.3}$$

$V(\hat{x}')$ はクラス $\hat{x}'$，$\hat{x}' \in C$ のキー情報ベクトルで，$F\left((\hat{x}', key_i) \middle| (x, y^N)^n \right)$ は学習データ全

体中でクラス $\hat{x}'$ に分類されている電報でキー情報 key_i が生起した回数，$A(key_i)$ は $F\left((\hat{x}', key_i)\middle|(x, y^N)^n\right) > 0$ が成立しているクラスの数，$V(doc_u)$ は新規に分類したい未知の電報 doc_u のキー情報ベクトルで，$F\left(key_i \middle| y'^{N'}\right)$ は未知の電報 doc_u のキー情報系列 $y'^{N'}$ 中でキー情報 key_i が生起した回数，cos はベクトル間の余弦の値を求める関数，$V(\hat{x}')\cdot V(doc_u)$ はベクトル $V(\hat{x}')$，$V(doc_u)$ 間の内積，$\|V(\hat{x}')\|$ はベクトル $V(\hat{x}')$ のノルムを示す．

この方法は距離を用いた分類方法の一つであり直感的に大変分り易く，言語的な性質やさまざまな経験則での微調整を行い易く，かつ計算量が小さいという長所があり，実用面での評価が非常に高い．しかし，その分類精度には理論的な保証はない．

文書分類問題の研究分野において，確率モデルを用いた様々な文書分類方法が提案されている[11][12][40]が，ここではその中でもその他の分野においても最もよく利用されている方法の一つである最尤推定法を採用した Naive-Bayes 法[8][19]（以下では単に Naive-Bayes 法と呼ぶ．）を取り上げる．Naive-Bayes 法に基づく文書分類方法を紹介する前に，いくつかの定義を行う．

$p(c_i|\theta)$ はクラス c_i が生起する確率分布，$p(key_j|c_i,\theta)$ はクラス c_i が生起した条件のもとでキー情報 key_j が生起する確率分布を示し，$p(c_i|\theta)$ も $p(key_j|c_i,\theta)$ もともに連続パラメータ θ，$\theta \in \Theta$ によって支配されている．θ^*，$\theta^* \in \Theta$ は真のパラメータで未知である

上記の確率モデルのもとでは，電報の生成とは，まず確率分布 $p(c_i|\theta)$ に従ってクラス c_i が生起し，次にいくつかのキー情報（キーワード）が独立に確率分布 $p(key_j|c_i,\theta)$ に従って生起することに相当する．パラメータ θ のもとでの学習用の電報 doc_{l_i} の生起確率 $p(doc_{l_i}|\theta)$ は次式で示される．

$$p(doc_{l_i}|\theta) = p(x_i|\theta)p(y^{N_i}|x_i,\theta) = p(x_i|\theta)\prod_{j=1}^{N_i} p(y_{i,j}|x_i,\theta). \tag{4.4}$$

現実の世界では各キー情報は $p(key_k|key_j,c_i,\theta)$ という確率分布で示されるように，一つ前に生起したキー情報によって次に生起するキー情報の確率分布が左右されるマルコフ性等を有すかも知れないが，上式では各キー情報が他のキー情報とは独立に生起すると仮定されている．現実世界のこのようなモデル化の仕方が Naive-Bayes 法での基本的な仮定である．本研究でも確率モデルには Naive-Bayes 法と同様の仮定を置いている．

同様に未知の電報 doc_u の生起確率 $p(doc_u|\theta)$ は次式で示される．

$$p(doc_u|\theta) = p(x'|\theta)p(y'^{N'}|x',\theta) = p(x'|\theta)\prod_{i=1}^{N'} p(y'_i|x',\theta). \quad (4.5)$$

次に，Naive-Bayes 法に基づく文書分類方法では未知の電報を分類するクラスが次式によって決定される．なお，$d_{NB}(y'^{N'})$は未知の電報doc_uのキー情報系列$y'^{N'}$を引数にとり，Naive-Bayes 法に基づいて未知の電報を分類するクラスを決定する決定関数である．

$$d_{NB}(y'^{N'}) = \arg\max_{\hat{x}' \in C} \hat{p}(\hat{x}')\prod_{i=1}^{N'} \hat{p}(y'_i|\hat{x}'), \quad (4.6)$$

ただし，$\hat{p}(\hat{x}')$および$\hat{p}(y'_i|\hat{x}')$はそれぞれ$p(\hat{x}'|\theta^*)$および$p(y'_i|\hat{x}',\theta^*)$に対する最尤推定法による推定値を示す．

上記の Naive-Bayes 法におけるパラメータの推定値が真のパラメータと一致する場合には，上式によって分類を間違えてしまう確率である誤り率を最小にするという意味での最適性が理論的に保証される．[6] Naive-Bayes 法では最尤推定法による推定値を採用しているので，漸近的に真のパラメータ既知の場合の最適な分類方法に収束することが理論的に保証されている．また，学習データが有限のもとでも学習データの増加に伴い分類精度が高まることが予測される．しかし，学習データの数が有限の場合には分類精度に関する厳密な理論的保証はない．

そこで，本研究では第 3 章で述べた統計的決定理論を学習問題に応用した際の共通モデルを適用し，学習データが有限の場合に未知の電報を間違ったクラスに分類してしまう確率である誤り率をベイズ基準のもとで最小化するという意味で最適な提案アルゴリズムを第 5 章で提案する．

４．１．３．既存名詞シソーラスへの未知語分類問題における従来研究

シソーラスとは，単語を意味的に分類した分類体系であり，あるまとまった単位の意味（または概念とも呼ばれる．）がクラスに相当する．シソーラスの多くは各クラスが各ノードに相当するようなツリー構造のクラス集合を有し，名詞の集合を分類した名詞シソーラスや，用言の集合を分類した用言シソーラスなどがある．また，木構造の葉のみにクラスが存在する分類シソーラスと，根及び中間ノードにもクラスが存在する上位下位シソーラスがある[26][30].

本研究では，クラス集合が木構造を有す名詞シソーラスを研究対象とし，分類シソーラスと上位下位シソーラスの区別は特に行わない．以下では，クラス集合が木構造を有す名詞シソーラスのことを単にシソーラスと呼ぶ．なお，シソーラス上の各クラスに分類されている名詞には一つのクラスのみに分類されている名詞と，複数のクラスに重複して分類されている名詞がある．各クラスにはそのクラスに分類されている名詞の集合の意味（または概念）を示すラベルが付与されている．章末の図 4.1 に上位下位シソーラスである NTT

シソーラス[10]の木構造の一部を示す．クラスを示す四角の中に付与されているのが概念で，各クラスの脇に付与されている番号は各クラスまたは各概念に対応する通し番号である．NTT シソーラスにおいて図 4.1 中のクラスに分類されている名詞の一部を章末の図 4.2 に示す．

本研究で扱う未知語分類問題とは，未知語を既存名詞シソーラス上のいずれかのクラスに分類する問題である．なお，ここで述べる未知語とは，既存名詞シソーラスには分類されていないが名詞であることが分かっている単語である．

未知語分類問題を考える場合には，ほとんどの従来研究がそうであるように，“ 単語の意味は，どのような単語と共起するかという観点から特徴づけられる ”という Harris の分布仮説[9]に基づいて考えるのが一般的である．そこで，本研究でも分布仮説に基づいて未知語分類問題を考えることにする．なお，二つの単語の共起の定義としては，その二つの単語が同一文中に共に存在すればいいという文内共起や，文の意味内容まで見て係り受け関係が成立しているもののみ共起とみなすものなどいろいろあるが，本研究では特に定めない．また，本研究では，名詞と動詞の共起のみを考慮する．よって，本研究では動詞がキー情報に相当する．

分布仮説に基づくと，シソーラスの同一クラスに分類されている名詞は動詞（キー情報）との共起の仕方が似ていると考えられる．すなわち，名詞と動詞の共起の仕方はシソーラスのクラスの数だけ存在し，未知語の共起の仕方もその中の一つであると考えられる．しかし，未知語の共起の仕方がその内のどれであるかは未知である．そこで，各クラスごとに動詞との共起の仕方を学習して，未知語の動詞との共起の仕方と比較し，未知語の共起の仕方と最も近い共起の仕方をするクラスに未知語を分類することによって，未知語分類問題を解決できる．ここで最も重要なことは，いかにして共起の仕方を比較するかである．

シソーラスへの未知語分類方法にはいろいろな従来研究があるが，自然言語処理の他の分野においても最もよく研究および実用化されている方法として，距離を用いた分類方法の代表的な方法であるベクトル空間法に基づく未知語分類方法と，確率モデルを用いた分類方法の代表的な方法である Naive-Bayes 法に基づく未知語分類方法を以下で説明する．

ベクトル空間法に基づく従来方法[41]は，シソーラスの各クラスのキー情報ベクトルと未知語のキー情報ベクトルの類似度をベクトル間の余弦を用いて算出し，類似度の高いクラスに未知語を分類するベクトル空間法を基本としている．最も単純なベクトル空間法では，キー情報ベクトルは名詞とキー情報（動詞）の共起頻度によるベクトルである．各クラスのキー情報ベクトルの各要素は，そのクラスに分類されている名詞のキー情報である動詞との共起頻度を足し合わせたもので，未知語のキー情報ベクトルの各要素は，未知語とキー情報の共起頻度である．実際には，単にベクトル空間法を用いるだけではなく，様々な言語的な性質を加味することによって分類精度を向上させる工夫がなされるが，本研究では従来研究のベクトル空間法に相当する部分が未知語分類問題の最も重要な部分であると考える．そこで，従来方法として基本的なベクトル空間法を以下で詳しく説明する．

まず，いくつかの定義を行う．$noun_i$ はシソーラス上のクラスに既に分類されている名詞を示し，$NOUN$，$NOUN = \{noun_1, noun_2, \cdots, noun_{|NOUN|}\}$ は要素数が有限の名詞集合である．なお，$|\cdot|$ は集合の要素数を示す．c_i はシソーラス上のクラスを示す．C，$C = \{c_1, c_2, \cdots, c_{|C|}\}$ は要素数が有限のシソーラス上のクラスの全集合で木構造を形成している．*unknown* は未知語を示す．key_i はキー情報を示し，共起を考慮する動詞に相当する．KEY，$KEY = \{key_1, key_2, \cdots, key_{|KEY|}\}$ は要素数が有限のキー情報の集合であり，動詞の集合に相当する．(x, y)，$x \in C$，$y \in KEY$ は一つの学習データを示す 2 項組であり，クラス x に分類されているいずれかの名詞とキー情報 y が示す動詞が共起したことを示す．$(x, y)^n$ は n 個の学習データからなる系列であり，$x^n y^n$，$x_1 y_1 x_2 y_2 \cdots x_n y_n$ と表記することもある．なお，学習データを生成するために用いる元々の文章のデータの中では名詞 $noun_i$ とキー情報（動詞）key_j が共起しているが，学習データを生成する時点で名詞とキー情報の 2 項組 $(noun_i, key_j)$ をクラスとキー情報の 2 項組 (c_k, key_j) に変換する．なお，クラス c_k は名詞 $noun_i$ が分類されているクラスであり，名詞 $noun_i$ が複数のクラスに分類されている場合は複数の 2 項組に変換する．$(x', y'^{N'})$，$x' \in C$，$y' \in KEY$ は未知語 *unknown* が分類されるべき真のクラス x' と未知語 *unknown* と共起した N' 個のキー情報 y' の系列 $y'^{N'}$ の 2 項組を示す．しかし，x' は未知であり，実際に与えられる未知語データは未知語 *unknown* と未知語 *unknown* と共起したキー情報 y' の系列 $y'^{N'}$ の 2 項組 $(unknown, y'^{N'})$ である．すなわち，未知語分類問題とは，学習データ $(x, y)^n$ と未知語データ $(unknown, y'^{N'})$ を与えられたもとで未知語 *unknown* の分類されるべきクラス x' を推定する問題である．

ベクトル空間法に基づく従来方法では，次式によって未知語を分類するクラスが決定される．

$$\begin{aligned} d_{Vec}\left((x, y)^n, (unknown, y'^{N'})\right) &= \arg\max_{\hat{x}' \in C} \cos\left(V(\hat{x}'), V(unknown)\right) \\ &= \arg\max_{\hat{x}' \in C} \frac{V(\hat{x}') \cdot V(unknown)}{\|V(\hat{x}')\| \|V(unknown)\|}, \end{aligned} \tag{4.7}$$

ただし，

$$V(\hat{x}') = \left(F\left((\hat{x}', key_1) \middle| (x, y)^n\right), F\left((\hat{x}', key_2) \middle| (x, y)^n\right), \cdots, F\left((\hat{x}', key_{|KEY|}) \middle| (x, y)^n\right)\right), \tag{4.8}$$

$$V(unknown) = \left(F\left(key_1 \middle| y'^{N'}\right), F\left(key_2 \middle| y'^{N'}\right), \cdots, F\left(key_{|KEY|} \middle| y'^{N'}\right)\right), \quad (4.9)$$

$d_{Vec}\left((x,y)^n, \left(unknown, y'^{N'}\right)\right)$は学習データ$(x,y)^n$と未知語データ$\left(unknown, y'^{N'}\right)$を引数にとり，未知語*unknown*を分類すべきクラスを決定する関数を示し，$V(\hat{x}')$はクラス$\hat{x}'$のキー情報ベクトル，$F\left(\left(\hat{x}', key_j\right) \middle| (x,y)^n\right)$は学習データ$(x,y)^n$中の$\left(\hat{x}', key_j\right)$の数でクラス$\hat{x}'$とキー情報$key_j$が共起した回数を示し，$V(unknown)$は未知語*unknown*のキー情報ベクトル，$F\left(key_i \middle| y'^{N'}\right)$は未知語データ$\left(unknown, y'^{N'}\right)$のキー情報系列$y'^{N'}$中のキー情報$key_i$の数で未知語*unknown*とキー情報$key_i$が共起した回数を示し，cos はベクトル間の余弦の値を求める関数，$V_A \cdot V_B$はベクトルV_A，V_B間の内積，$\|V\|$はベクトルVのノルムを示す．

式(4.7)で示されるように，従来方法では未知語のキー情報ベクトル$V(unknown)$との余弦の値が最大になるキー情報ベクトル$V(\hat{x}')$に対応するクラス$\hat{x}'$に未知語*unknown*を分類する．章末の図 4.3 で説明すると，ベクトル空間において，各ベクトルの長さは無視して未知語*unknown*のキー情報ベクトルと成す角度が最小のキー情報ベクトルに対応するクラス$\hat{x}'$に未知語*unknown*を分類する．これは，未知語*unknown*をキー情報との共起の仕方の相関が最大となるクラス$\hat{x}'$に分類していると解釈できる[36][37]．この方法は距離を用いた分類方法の１種であり直感的に大変分り易く，言語的な性質やさまざまな経験則での微調整を行い易く，かつ計算量が小さいという長所があり，実用面での評価が非常に高い．しかし，その分類精度には理論的な保証はない．

上記のベクトル空間法の微調整の一例として TF･IDF 法が挙げられる．これは各共起頻度に重み付けを行う方法であり，情報検索等の分野において多く実用化されているのは TF･IDF 法を導入したベクトル空間法である[14]．TF･IDF 法を導入したベクトル空間法では，式(4.8)および式(4.9)のキー情報ベクトルの第i要素に$\log \frac{|C|}{A(key_i)}$を掛け合わせたものをキー情報ベクトルとして採用し，式(4.7)で未知語の分類を行う．ただし，$A(key_i)$はキー情報key_iとの共起頻度が1以上のクラスの数である．未知語分類問題に限らず一般的に単なる共起頻度によるベクトル空間法よりも，TF･IDF 法を導入したベクトル空間法の方が精度が高いことを示す事例が数多く報告されており，情報検索等いろいろな分野で実用化されている．

次に Naive-Bayes 法に基づく未知語分類方法に関しては，必ずしも Naive-Bayes 法をシソーラスへの未知語分類問題に適用した従来研究があるわけではないが，Naive-Bayes 法は自然言語処理の他の分類問題においては確率モデルを用いた分類方法の中で代表的な方

法として最もよく研究されている方法であるので, ここでは未知語分類問題に Naive-Bayes 法を適用した場合を想定して，Naive-Bayes 法に基づく未知語分類方法について以下で説明する．[8]

まず，いくつかの定義を行う．クラス c_i の生起する確率分布を $p(c_i|\theta)$，クラス c_i が生起したもとでキー情報 key_j が生起する確率分布を $p(key_j|c_i,\theta)$ と表す． $p(c_i|\theta)$ と $p(key_j|c_i,\theta)$ はともに連続パラメータ θ によって支配され，パラメータ集合を Θ とし，真のパラメータ θ^*， $\theta^* \in \Theta$ は未知とする．

Naive-Bayes 法を未知語分類問題に適用すると，次式のような未知語分類方法 $d_{NB}(y'^{N'})$ が考えられる．

$$d_{NB}(y'^{N'}) = \arg\max_{\hat{x}' \in C} \hat{p}(\hat{x}') \prod_{j=1}^{N'} \hat{p}(y'_j|\hat{x}'), \tag{4.10}$$

ただし， $\hat{p}(\hat{x}')$ は $p(\hat{x}'|\theta^*)$ の最尤推定法による推定値， $\hat{p}(y'_j|\hat{x}')$ は $p(y'_j|\hat{x}',\theta^*)$ の最尤推定法による推定値を示す．式(4.10)による未知語分類方法は推定値を真のパラメータと仮定して, 真のパラメータ既知の場合に誤り率を最小にする方法に推定値を代入している．パラメータの推定値が真のパラメータと一致する場合には，上式によって分類先を間違えてしまう確率である誤り率を最小にするという意味での最適性が理論的に保証されている．上式では最尤推定法による推定値を採用しているので，漸近的に真のパラメータ既知の場合の最適な分類方法に収束することが理論的に保証されている．また，学習データが有限のもとでも学習データの増加に伴い分類精度が高まることが予測される．しかし，学習データの数が有限の場合には分類精度に関する厳密な理論的保証はない.

そこで，本研究では第 3 章で述べた統計的決定理論を学習問題に応用した際の共通モデルを適用し，学習データが有限の場合に未知語を間違ったクラスに分類してしまう確率である誤り率をベイズ基準のもとで最小化するという意味で最適な提案アルゴリズムを第 5 章で提案する.

４．２．マルコフ決定過程問題における従来研究[13][21][39]

４．２．１．マルコフ決定過程問題の概要

マルコフ決定過程(MDP， Markov Decision Processes)問題は有限の状態, 有限の行動, 遷移確率行列，そして遷移に伴う収益を示す利得関数によって定義される．マルコフ決定過程問題における目的は，行動を選択し，状態が遷移し，その状態遷移に伴う収益を得るという一連のプロセスを繰り返すことによって得られる総収益を最大化することである.

選択するべき行動を決定する決定関数を政策と呼び，総収益を最大化するような政策を最適政策と呼ぶ.

まず，いくつかの定義を行う．遷移確率行列を支配する連続パラメータをθと表し，そのパラメータ集合をΘ, 真のパラメータをθ^*, $\theta^* \in \Theta$と表す. 有限の状態集合をS, $s_i \in S$, 有限の行動集合をA，$a_k \in A$，状態s_iにおいて行動a_kが選択されたもとで状態s_jに遷移したときに得られる収益を示す利得関数を$r\left(s_i, a_k, s_j\right)$と表す．状態$s_i$において行動$a_k$が選択されたもとで，状態$s_j$に遷移する確率を示す，パラメータ$\theta$によって支配される$|S||A| \times |S|$の遷移確率行列($|\cdot|$は集合の濃度を示す．)の要素を$p\left(s_j | s_i, a_k, \theta\right)$と表す．行動選択と状態遷移を繰り返して収益を得る期間である運用期間の長さをT，運用期間の初期状態をx_0，t期における状態をx_t，t期において選択された行動をd_tと表し，t期における状態遷移に伴って得ることができた収益をU'_t，状態x_0において行動d_0が選択され，状態x_1へ遷移し，状態x_1において行動d_1が選択され，という一連の遷移の履歴を示す遷移系列を$x_0 d_0 x_1 d_1 \cdots x_t$などと表す.

マルコフ決定過程問題には，割引問題と非割引問題という二種類の問題がある．割引問題における目的は，次式で示される期待割引総収益の最大化である．（厳密にはその目的はさらにさまざまな種類に分かれるが，それについては後述する．）

$$v = E\left(\sum_{t=0}^{T} \beta^t U'_t\right), \tag{4.11}$$

ただし，β，$0 < \beta < 1$は割引率である.

非割引問題における目的は次式で示される期待平均収益の最大化である.

$$h = E\left(\lim_{T\to\infty}\left(\sum_{t=0}^{T} U'_t \Big/ T+1\right)\right). \tag{4.12}$$

本研究では，マルコフ決定過程問題の中でも特に割引問題を扱うこととし，以下ではマルコフ決定過程問題の中の割引問題を単にマルコフ決定過程問題と呼ぶこととする．以下でマルコフ決定過程問題について詳述する.

マルコフ決定過程問題はすべての情報が既知の問題と未知情報を伴う問題とに大別され，本研究では未知情報を伴う問題を研究対象とするが，まずここでは，すべての情報が既知の場合について述べる．運用期間Tが有限の固定期間の場合には，動的計画法(DP, Dynamic Programming)[4][13][21][39]で次式の再帰方程式を解くことによって，式(4.11)の期待割引総収益を最大化する最適政策が求められる.

$$v^*(x_t, T-t) = \max_{d_t} \sum_{x_{t+1}} p(x_{t+1}|x_t, d_t, \theta^*)(r(x_t, d_t, x_{t+1}) + \beta v^*(x_{t+1}, T-(t+1))), \quad (4.13)$$

ただし, $v^*(x_t, T-t)$は真のパラメータ既知のもとで状態x_tを初期状態とした$T-t$期間の真の最適政策による期待割引総収益を示す．ここで真の最適政策とは真のパラメータ既知の場合の最適政策のことである．また，運用期間Tが無限の場合には，Policy Iteration Algorithm(PIA)[13][21][39]で次式の再帰方程式を解くことによって，式(4.11)の期待割引総収益を最大化する決定関数である最適政策が求められる．

$$v^*(s_i) = \max_{a_k} \sum_{s_j} p(s_j|s_i, a_k, \theta^*)(r(s_i, a_k, s_j) + \beta v^*(s_j)), \quad (4.14)$$

ただし, $v^*(s_i)$は真のパラメータ既知のもとで状態s_iを初期状態とした無限期間の真の最適政策による期待割引総収益を示す.

４．２．２．未知情報を伴うマルコフ決定過程問題における従来研究

未知情報を伴うマルコフ決定過程問題では，未知情報について学習する必要がある．従来研究の中には，漸近的に真の最適政策への収束を保証することを目的とした研究にQ-Learning[42]がある．Q-Learningでは運用期間が無限の未知情報を伴うマルコフ決定過程問題に対して，真の最適政策を求めることを目的とし，次式でQ値を更新することにより未知情報について学習すると共に，なるべく大きな収益を得るような行動選択を行う．

$$Q_{t+1}(x_t, d_t) = (1-\alpha)Q_t(x_t, d_t) + \alpha\left(r(x_t, d_t, x_{t+1}) + \beta \max_{d_{t+1} \in A} Q_t(x_{t+1}, d_{t+1})\right), \quad (4.15)$$

ただし，α，$0 < \alpha < 1$は学習率，$Q_t(s_i, a_k)$は$\sum_{s_j} p(s_j|s_i, a_k, \theta^*)(r(s_i, a_k, s_j) + \beta v^*(s_j))$の$t$期における近似値を示す．Q-Learningは確率的近似法と同様に収束性が保証されており，漸近的に$Q_t(s_i, a_k)$が$\sum_{s_j} p(s_j|s_i, a_k, \theta^*)(r(s_i, a_k, s_j) + \beta v^*(s_j))$に収束する．よって，Q値が最大となる行動を選択することを続けることによって式(4.14)による真の最適政策を漸近的に求めることができる．Q-Learningは実装が容易であり，かつ計算量も少ないので多くの分野で適用例が報告されている．真の最適政策への収束が保証されているので，例えばロボット制御の場合，実験環境において十分に長い運用期間を設けて良好な政策を得た後に，その政策を実環境で使用するロボットに設定するという利用の仕方も出来，大変有効なアルゴリズムである．しかし，有限の運用期間における収益に関する理論的な保証はない．

一方，有限の運用期間における収益の最大化を目的とした研究にMartinの研究[20][34]がある．Martinの研究では運用期間が有限の固定期間の問題に対して本研究と同様に統計的決定理論[3][7][35]を応用し，DPの問題を解くことによって，運用期間の期待割引総収益

をベイズ基準のもとで最大化している．Martin のアルゴリズムについて説明する前にまずいくつか定義を追加する．

パラメータ θ の事前確率密度関数を $p(\theta)$，$x_0 d_0 x_1 d_1 \cdots x_t$ という遷移をしたもとでの事後確率密度関数を $p(\theta | x_0 d_0 x_1 d_1 \cdots x_t)$ と表す．次式によって $x_0 d_0 \cdots x_t$ という遷移をしたもとで，状態 x_t において行動 d_t を選択して状態 x_{t+1} へ遷移する確率の，遷移確率行列を支配するパラメータの事後確率密度関数による期待値を定義する．

$$\bar{p}(x_{t+1} | x_t, d_t, x_0 d_0 \cdots x_t) = \int_{\Theta} p(\theta | x_0 d_0 \cdots x_t) p(x_{t+1} | x_t, d_t, \theta) d\theta. \tag{4.16}$$

Martin のアルゴリズムでは次式の再帰方程式を解くことによって，その期以後の期待割引総収益をベイズ基準のもとで最大化する行動が決定される．

$$\begin{aligned} & d_t(x_0 d_0 \cdots x_t) \\ & \quad = \arg\max_{d_t} \sum_{x_{t+1}} \bar{p}(x_{t+1} | x_t, d_t, x_0 d_0 \cdots x_t)(r(x_t, d_t, x_{t+1}) + \beta U''(x_0 d_0 \cdots x_t d_t x_{t+1})), \end{aligned} \tag{4.17}$$

ただし，
$d_t(x_0 d_0 \cdots x_t)$ は $x_0 d_0 \cdots x_t$ という遷移をしたもとで選択するべき行動を決定する決定関数を示し，$U''(x_0 d_0 \cdots x_{t+1})$ は以下の式のように $x_0 d_0 \cdots x_{t+1}$ という遷移をしたもとでの，それ以降の期待割引総収益のベイズ基準のもとでの最大値を示し，

$$\begin{aligned} & U''(x_0 d_0 \cdots x_{t+1}) \\ & \quad = \max_{d_{t+1}} \sum_{x_{t+2}} \bar{p}(x_{t+2} | x_{t+1}, d_{t+1}, x_0 d_0 \cdots x_{t+1})(r(x_{t+1}, d_{t+1}, x_{t+2}) + \beta U''(x_0 d_0 \cdots x_{t+1} d_{t+1} x_{t+2})), \end{aligned} \tag{4.18}$$

かつ

$$U''(x_0 d_0 \cdots x_T) = 0. \tag{4.19}$$

以上のように，Martin のアルゴリズムでは，事後確率を更新することによって未知情報について学習すると共に，事後確率で期待値を算出した期待割引総収益を最大にするように行動選択を行うことによって，統計的決定理論に基づきベイズ基準のもとで期待割引総収益の最大化を実現している．

Q-Learning や Martin のアルゴリズムでは，当該期間に得られる収益を評価対象とする運用期間のみからなる未知情報を伴うマルコフ決定過程問題を扱っている．他方，別の問題設定として，当該期間の状態遷移によって発生する収益は無視して未知情報の学習に専念する準備期間が運用期間の前にある問題設定も考えられる．この様な問題を扱った従来研究には宮崎等[27][28][29]や Barto 等[2]の研究がある．これらの従来研究では，準備期間と運用期間に分割された未知情報を伴うマルコフ決定過程問題に対して，準備期間にはマルコフ決定過程問題の遷移確率行列を支配する未知パラメータの最尤推定が行われ，運用期間には準備期間で求めたパラメータの推定値を真のものと仮定して PIA（運用期間の長

さが無限の場合）または DP（運用期間の長さが有限の場合）で求めた政策を用いて制御が行われる．これらの研究では，準備期間と運用期間を明確に分割して，準備期間は未知パラメータの学習に専念し，運用期間には準備期間に得られた情報を駆使して期待割引総収益を最大化しようとしている．パラメータの推定方法として最尤推定法を採用しているため，準備期間の長さが長くなれば漸近的に運用期間の政策が真の最適政策に収束することが保証されている．しかし，有限の準備期間のもとでは推定誤差があるため，期待割引総収益最大化の厳密な保証はない．

そこで，本研究では有限で固定の準備期間と有限で固定の運用期間の問題に対して，運用期間の期待割引総収益に何らかの保証を与えることを目的として，第 3 章で述べた統計的決定理論を学習問題に応用した際の共通モデルを適用し，運用期間の期待割引総収益をベイズ基準のもとで最大化する提案アルゴリズムを第 6 章で提案する．なお，本研究は Martin の研究の問題設定を拡張した研究として解釈できる．

４．３．本研究の位置付け

章末の図 4.4 および図 4.5 に本研究の位置付けを示す．文書分類問題，既存名詞シソーラスへの未知語分類問題，準備期間と運用期間に分割された未知情報を伴うマルコフ決定過程問題に関して，従来研究においては実装の容易さや言語的性質の加味等の工夫のし易さを重要視するような研究や，学習データが豊富な場合に漸近的に真のパラメータ既知の場合の真の最適解への収束を保証するような従来方法が提案されている．本研究ではこれらの問題に対して，第 3 章で述べた統計的決定理論を学習問題に応用した際の共通モデルを適用し，従来研究では未だ研究されていない，学習データが有限の場合にベイズ基準のもとで文書分類問題等における誤り率を最小化する提案アルゴリズムや，準備期間と運用期間に分割された未知情報を伴うマルコフ決定過程問題における期待割引総収益を最大化する提案アルゴリズムを導出する．

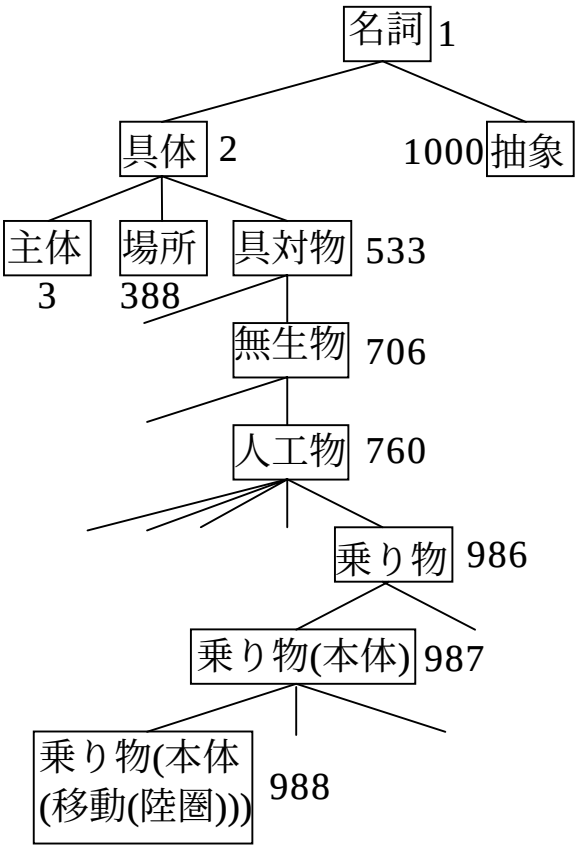

図 4.1　NTT シソーラスのクラス集合の木構造

986　乗り物

空便　初便　先便　増便

定期便　便　夜行便

987　乗り物(本体)

乗りもの　乗り物　乗物

988　乗り物(本体(移動(陸圏)))

愛車　愛用車　青電車

図 4.2　NTT シソーラスの各クラスに分類されている名詞

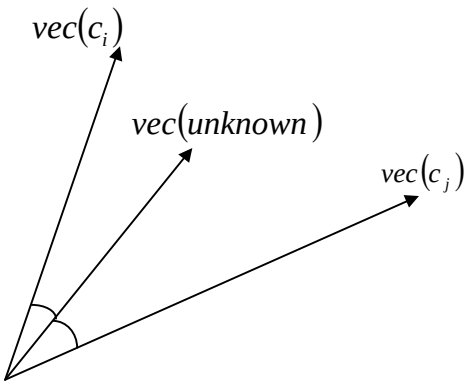

図 4.3　ベクトル空間における未知語分類

		距離を用いた方法	確率モデルを用いた方法	
		実装上の利便性を重視	真の最適解を同定	ベイズ基準のもとで誤り率最小化
分類問題	文書分類問題	従来研究 第4章	従来研究 第4章	本研究 第5章
	既存名詞シソーラスへの未知語分類問題	従来研究 第4章	従来研究 第4章	本研究 第5章

図 4.4　分類問題に関する本研究の位置付け

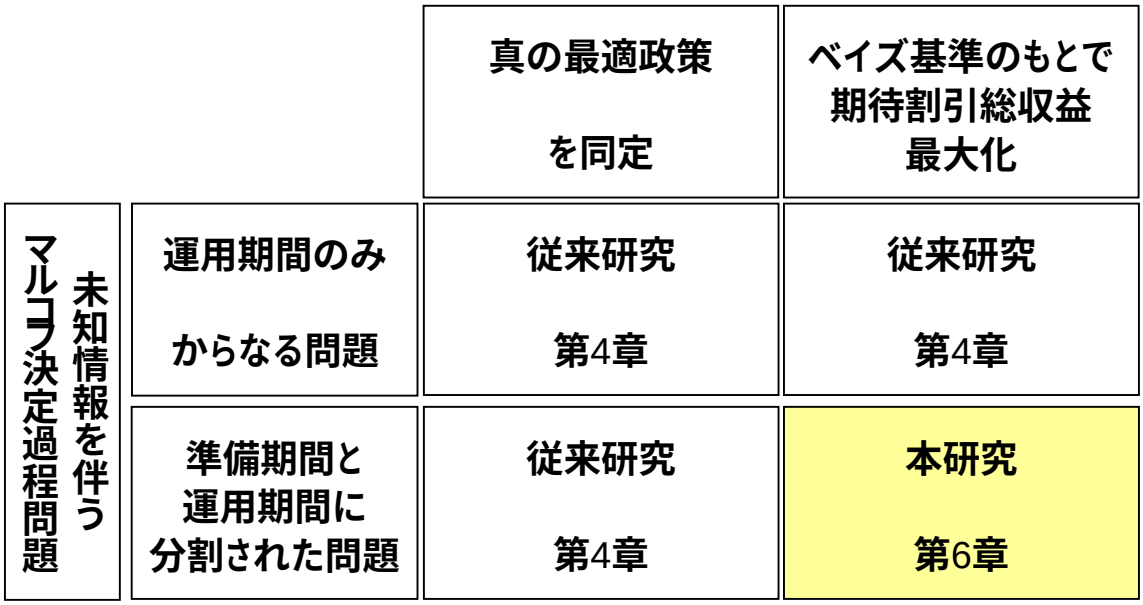

図 4.5　未知情報を伴うマルコフ決定過程問題に関する本研究の位置付け